JOURNAL OF MULTIVARIATE ANALYSIS 5, 248–264 (1975)

Reduced-Rank Regression for the Multivariate Linear Model

ALAN JULIAN IZENMAN

Department of Statistics, Tel Aviv University, Israel

Communicated by P. R. Krishnaiah

The problem of estimating the regression coefficient matrix having known (reduced) rank for the multivariate linear model when both sets of variates are jointly stochastic is discussed. We show that this problem is related to the problem of deciding how many principal components or pairs of canonical variates to use in any practical situation. Under the assumption of joint normality of the two sets of variates, we give the asymptotic (large-sample) distributions of the various estimated reduced-rank regression coefficient matrices that are of interest. Approximate confidence bounds on the elements of these matrices are then suggested using either the appropriate asymptotic expressions or the jackknife technique.

1. INTRODUCTION

This paper is concerned with investigating an important generalization of the well-known technique of multiple regression analysis, of which much has been written for both uniresponse and multiresponse situations. The aspect of this subject that will be of interest here is the following. Consider the general multivariate linear model,

$$\overset{s\times n}{\mathbf{Y}} = \overset{s\times n}{\boldsymbol{\mu}} + \overset{s\times r}{\mathbf{C}}\ \overset{r\times n}{\mathbf{X}} + \overset{s\times n}{\mathscr{E}}, \tag{1.1}$$

where the n columns of $\mathbf{X}$ and $\mathbf{Y}$ are multiresponse observations, $\boldsymbol{\mu}$ and $\mathbf{C}$ are unknown parameters, and $\mathscr{E}$ is the matrix of residuals. Under certain assumptions on $\mathbf{X}$, $\mathbf{Y}$, and $\mathscr{E}$, the problem is to estimate and provide confidence bounds for $\boldsymbol{\mu}$ and $\mathbf{C}$, where we impose the extra restriction that the regression coefficient matrix $\mathbf{C}$ has a certain (known or unknown) rank. The usual descriptions of this

Received March 6, 1974; revised December 2, 1974.

AMS 1970 subject classification: Primary H20, H25.

Key words and phrases: Multivariate linear regression, principal components, canonical variates, asymptotic distribution theory, confidence bounds, jackknife.

model (1.1) assume implicitly that $\mathbf{C}$ has full-rank, and then demonstrate that simultaneous least-squares estimation applied to all the equations of the model yields the same results as does equation-by-equation least-squares. Therefore, nothing is gained by estimating the equations jointly. A new feature enters the multivariate case when we admit the possibility that the regression coefficient matrix may not have full-rank; i.e., there may exist linear restrictions on the regression coefficients.

The main results concerning the estimation of a regression coefficient matrix of known reduced-rank were obtained by Rao [14, p. 505] and by Brillinger [3] and are restated below. For the case where the rank of $\mathbf{C}$ is unknown, a method of estimating this rank is indicated in Izenman [6]; then, using this estimate as the true rank, an appropriate estimate of $\mathbf{C}$ can subsequently be obtained. Anderson [1] obtained the likelihood-ratio test of the hypothesis that the rank of $\mathbf{C}$ is a given number and carried out the associated asymptotic theory under the assumption of normality of the variables. See also Anderson [2, Sect. 14.2] and, more recently, Fujikoshi [4]. Robinson [16] considered a similar problem within a time-series framework. A further reference is [18].

A by-product of this approach to multivariate multiple regression analysis is its application to principal component analysis and canonical variate and correlation analysis. The importance of these latter techniques lies in their ability (or inability) to perform a linear reduction of dimensionality of a given multivariate data set. The effective number of principal components or pairs of canonical variates is determined by the maximal amount of reduction in dimensionality of the data that is possible without destroying the informational content. We show below that, under certain conditions, there is a correspondence between the rank of the regression coefficient matrix $\mathbf{C}$ in (1.1) and a specified number of principal components or pairs of canonical variates.

In this paper we develop the asymptotic theory necessary to provide approximate (large-sample) confidence bounds for the matrix $\mathbf{C}$ in (1.1) having known reduced-rank. For the case corresponding to the principal components situation, explicit computable bounds can be obtained; however, for the canonical variates case, this does not appear feasible and an alternative method (the jackknife) is suggested in its place.

2. Reduced-Rank Regression When X and Y are Jointly Stochastic

Let

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \tag{2.1}$$

be an $(r + s)$-vector variate, with $\mathbf{X}$ r-vector-valued and $\mathbf{Y}$ s-vector-valued, where we shall assume that $s \leqslant r$. The variate (2.1) is assumed to come equipped with a mean vector and a covariance matrix given by

$$\begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{bmatrix}, \tag{2.2}$$

respectively. We shall assume also that the variates $\mathbf{X}$ and $\mathbf{Y}$ are related in the following linear fashion,

$$\overset{s\times 1}{\mathbf{Y}} = \overset{s\times 1}{\boldsymbol{\mu}} + \overset{s\times r}{\mathbf{C}} \overset{r\times 1}{\mathbf{X}} + \overset{s\times 1}{\mathscr{E}}, \tag{2.3}$$

where $\boldsymbol{\mu}$ and $\mathbf{C}$ are unknown parameters, $\mathscr{E}$ is the corresponding error-variate of the model, and the regression coefficient matrix $\mathbf{C}$ has a specified rank t, say, where $1 \leqslant t \leqslant s$. The usual case is where $t = s$. To distinguish it from the case $1 \leqslant t < s$, the former will be called the *full-rank regression coefficient matrix* and the latter the *reduced-rank regression coefficient matrix*. In the event that $1 \leqslant t < s$, there exists an $(s \times t)$-matrix $\mathbf{A}$ and a $(t \times r)$-matrix $\mathbf{B}$, each having rank t, such that $\mathbf{C} = \mathbf{AB}$. The $\mathbf{A}$ and $\mathbf{B}$ are not unique; indeed, if $\mathbf{T}$ is any $(t \times t)$-nonsingular matrix, then $\mathbf{C} = (\mathbf{AT})(\mathbf{T}^{-1}\mathbf{B}) = \mathbf{DE}$ is also a decomposition of $\mathbf{C}$. Thus, we wish to determine $\boldsymbol{\mu}$, $\mathbf{A}$ and $\mathbf{B}$ to minimise the error-variate $\mathbf{Y} - \boldsymbol{\mu} - \mathbf{ABX}$. We have the following theorem; for a proof, see Brillinger [3, Theor. 10.2.1] or Izenman [5, Theor. 2.1.1.].

THEOREM 1. *Let* (2.1) *be an* $(r + s)$*-vector-valued variate having mean vector and covariance matrix given by* (2.2) *respectively. Suppose that* $\boldsymbol{\Sigma}_{XX}$ *is nonsingular and* $\boldsymbol{\Gamma}$ *is positive-definite symmetric. Then, an* $(s \times 1)$*-vector* $\boldsymbol{\mu}$, *an* $(s \times t)$*-matrix* $\mathbf{A}$ *and a* $(t \times r)$*-matrix* $\mathbf{B}$, *where* $1 \leqslant t \leqslant s \leqslant r$, *that minimise simultaneously all the latent roots of*

$$E\{\boldsymbol{\Gamma}^{1/2}(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{ABX})(\mathbf{Y} - \boldsymbol{\mu} - \mathbf{ABX})^{\tau}\boldsymbol{\Gamma}^{1/2}\} \tag{2.4}$$

are given by

$$\mathbf{A} = \boldsymbol{\Gamma}^{-1/2}[\mathbf{V}_1, \dots, \mathbf{V}_t] \tag{2.5}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{V}_1^{\tau} \\ \vdots \\ \mathbf{V}_t^{\tau} \end{bmatrix} \boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}, \tag{2.6}$$

$$\boldsymbol{\mu} = \boldsymbol{\mu}_Y - \mathbf{AB}\boldsymbol{\mu}_X, \tag{2.7}$$

where $\mathbf{V}_j$ *is the* jth *latent vector of* $\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Gamma}^{1/2}$, $j = 1, 2, \dots, s$.

A possible interpretation of such a set-up is that of using the r components of the vector $\mathbf{X}$ to send a message $\mathbf{Y}$ having s components $(s \leqslant r)$, but where such a message can only be transmitted using t channels ($t \leqslant s$). Thus, the procedure would be to code the original $\mathbf{X}$ into a t-component vector $\mathbf{BX}$, and on receipt of this to form the s-component $\mathbf{ABX}$ which, it would be hoped, would be as close as possible to the desired $\mathbf{Y}$. If the criterion of "closeness" is given by (2.4), then the appropriate $\mathbf{A}$ and $\mathbf{B}$ are given by Theorem 1.

Setting $\mathbf{C} = \mathbf{AB}$, we see from Theorem 1 that the required regression coefficient matrix $\mathbf{C}$ having rank t is, therefore, given by the expression

$$\mathbf{C} = \mathbf{\Gamma}^{-1/2} \left(\sum_{j=1}^{t} \mathbf{V}_j \mathbf{V}_j^{\tau} \right) \mathbf{\Gamma}^{1/2} \mathbf{\Sigma}_{YX} \mathbf{\Sigma}_{XX}^{-1}, \tag{2.8}$$

where $\mathbf{V}_j$ is the jth latent vector of $\mathbf{\Gamma}^{1/2}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Gamma}^{1/2}$. When $t = s$, the matrix $\mathbf{C}$ in (2.8) reduces to the usual full-rank (least-squares) matrix of regression coefficients, to be denoted by $\boldsymbol{\theta} = \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}$. Thus, $\mathbf{C}$ and $\boldsymbol{\theta}$ are related by the equation $\mathbf{C} = \mathbf{P}_{\Gamma}\boldsymbol{\theta}$, where $\mathbf{P}_{\Gamma}$ is an idempotent matrix for any $\mathbf{\Gamma}$, but will generally not be symmetric. In the following sections of this paper, we shall be more specific in our choice of the matrix $\mathbf{\Gamma}$. These choices are motivated primarily by their relationships to the classical multivariate techniques of principal components and canonical variate analysis.

3. Relationship to Principal Components Analysis

Let $\mathbf{X}$ be an r-vector-valued variate having mean $\boldsymbol{\mu}_X$ and covariance matrix $\mathbf{\Sigma}_{XX}$. The problem is to find a $(t \times r)$-matrix $\mathbf{B}$, with $1 \leqslant t \leqslant r$, such that the variate $\boldsymbol{\zeta} = \mathbf{BX}$ minimises the loss in statistical information in going from $\mathbf{X}$ to $\boldsymbol{\zeta}$. A suitable criterion (see [11]) consists of determining an r-vector $\boldsymbol{\mu}$, an $(r \times t)$-matrix $\mathbf{A}$, and a $(t \times r)$-matrix $\mathbf{B}$ for which the latent roots of

$$E\{(\mathbf{X} - \boldsymbol{\mu} - \mathbf{ABX})(\mathbf{X} - \boldsymbol{\mu} - \mathbf{ABX})^{\tau}\} \tag{3.1}$$

are simultaneously minimized. This is clearly a special case of (2.4) with $\mathbf{\Gamma} = \mathbf{I}_s$, $\mathbf{Y} \equiv \mathbf{X}$ and $s = r$. Appropriate $\boldsymbol{\mu}$, $\mathbf{A}$ and $\mathbf{B}$ are, therefore, given by

$$\mathbf{A} = [\mathbf{V}_1, \ldots, \mathbf{V}_t] = \mathbf{B}^{\tau}, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_X - \mathbf{AB}\boldsymbol{\mu}_X, \tag{3.2}$$

where $\mathbf{V}_j$ is the jth latent vector of $\mathbf{\Sigma}_{XX}$, $j = 1, 2, \ldots, r$. The components of the vector $\boldsymbol{\zeta}$ are given by $\zeta_j = \mathbf{V}_j^{\tau}\mathbf{X}$, $j = 1, 2, \ldots, t$. If λ_j is the corresponding latent root, then $\mathrm{var}\{\zeta_j\} = \lambda_j$; and $\mathrm{corr}\{\zeta_j, \zeta_k\} = 0$ for $j \neq k$. These individual components of $\boldsymbol{\zeta}$ are commonly referred to as the *principal components of* $\mathbf{X}$.

Setting $\mathbf{C} = \mathbf{AB}$ in (3.1) shows that the principal components problem is equivalent to that of a reduced-rank regression, in the sense that the model can be written as

$$\underset{r\times 1}{\mathbf{X}} = \underset{r\times 1}{\boldsymbol{\mu}} + \underset{r\times r}{\mathbf{C}}\ \underset{r\times 1}{\mathbf{X}} + \underset{r\times 1}{\mathscr{E}}, \tag{3.3}$$

and where we wish to find the r-vector $\boldsymbol{\mu}$ and the $(r \times r)$-matrix $\mathbf{C}$ having reduced-rank t, $1 \leqslant t \leqslant r$, which minimise simultaneously all the latent roots of

$$E\{(\mathbf{X} - \boldsymbol{\mu} - \mathbf{CX})(\mathbf{X} - \boldsymbol{\mu} - \mathbf{CX})^{\tau}\}. \tag{3.4}$$

From (3.2), the appropriate $\boldsymbol{\mu}$ and $\mathbf{C}$ are given by

$$\mathbf{C} = \sum_{j=1}^{t} \mathbf{V}_j\mathbf{V}_j^{\tau}, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_X - \mathbf{C}\boldsymbol{\mu}_X. \tag{3.5}$$

We shall, henceforth, refer to the model (3.3) together with the minimisation criterion (3.4) (equivalent to setting $\boldsymbol{\Gamma} = \mathbf{I}_s$, $\mathbf{Y} \equiv \mathbf{X}$ and $r = s$ in the general set-up (2.3, 2.4)) as the *reduced-rank regression model corresponding to the principal components situation.* The solution (3.5) will be referred to similarly. When $t = r$, $\mathbf{C}$ in (3.5) becomes $\mathbf{I}_r$, the $(r \times r)$-identity matrix; hence the inclusion of a residual vector $\mathscr{E}$ in the model (3.3) becomes superfluous. However, this should not cause any problems in interpretation.

4. Relationship to Canonical Variate Analysis

In the classical case of canonical variate analysis, the purpose is to obtain a $(t \times r)$-matrix $\mathbf{G}$ and a $(t \times s)$-matrix $\mathbf{H}$, with $1 \leqslant t \leqslant s \leqslant r$, such that the new t-vector variates $\boldsymbol{\xi} = \mathbf{GX}$ and $\boldsymbol{\omega} = \mathbf{HY}$ will retain almost all of the statistical relationship between the r-vector variate $\mathbf{X}$ and the s-vector variate $\mathbf{Y}$. We have the following theorem; for a proof, see Brillinger [3, Theor. 10.2.2] or Izenman [5, Theor. 2.3.1].

Theorem 2. *Let* (2.1) *be an* $(r + s)$*-vector-valued variate having mean vector and covariance matrix given by* (2.2) *respectively, and suppose that both* $\boldsymbol{\Sigma}_{XX}$ *and* $\boldsymbol{\Sigma}_{YY}$ *are nonsingular. Then, the* $(t \times 1)$*-vector* $\boldsymbol{\nu}$, *the* $(t \times r)$*-matrix* $\mathbf{G}$ *and the* $(t \times s)$*-matrix* $\mathbf{H}$ *with* $\mathbf{H}\boldsymbol{\Sigma}_{YY}\mathbf{H}^{\tau} = \mathbf{I}_t$ *that minimise simultaneously all the latent roots of*

$$E\{(\mathbf{HY} - \boldsymbol{\nu} - \mathbf{GX})(\mathbf{HY} - \boldsymbol{\nu} - \mathbf{GX})^{\tau}\} \tag{4.1}$$

are given by

$$\mathbf{G} = \begin{bmatrix} \mathbf{V}_1^\tau \\ \vdots \\ \mathbf{V}_t^\tau \end{bmatrix} \mathbf{\Sigma}_{YY}^{-1/2}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}, \tag{4.2}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{V}_1^\tau \\ \vdots \\ \mathbf{V}_t^\tau \end{bmatrix} \mathbf{\Sigma}_{YY}^{-1/2}, \tag{4.3}$$

$$\boldsymbol{\nu} = \mathbf{H}\boldsymbol{\mu}_Y - \mathbf{G}\boldsymbol{\mu}_X, \tag{4.4}$$

where $\mathbf{V}_j$ *is the latent vector corresponding to the* jth *largest latent root* λ_j *of*

$$\mathbf{R} = \mathbf{\Sigma}_{YY}^{-1/2}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1/2}, \qquad j = 1, 2,\ldots, s. \tag{4.5}$$

Compare Theorem 2 with Theorem 1. When $\mathbf{\Gamma} = \mathbf{\Sigma}_{YY}^{-1}$, the matrix $\mathbf{B}$ in (2.6) and the matrix $\mathbf{G}$ in (4.2) are identical. Furthermore, the matrix $\mathbf{A}$ in (2.5) and the matrix $\mathbf{H}$ in (4.3) satisfy the relationships, $\mathbf{HAH} = \mathbf{H}$ and $\mathbf{AHA} = \mathbf{A}$. In the special case that $t = s$, the further relations $(\mathbf{AH})^\tau = \mathbf{AH}$ and $(\mathbf{HA})^\tau = \mathbf{HA}$ hold, and so $\mathbf{H} = \mathbf{A}^+$, the unique *Moore–Penrose generalized inverse* of $\mathbf{A}$ (cf. [15]). Computationally, then, the $\mathbf{G}$, $\mathbf{H}$ and $\boldsymbol{\nu}$ of Theorem 2 can be obtained directly from the $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\mu}$ of Theorem 1 (and, of course, vice versa). First, the matrices $\mathbf{A}$ in (2.5) and $\mathbf{B}$ in (2.6) are computed, and then $\mathbf{C} = \mathbf{AB}$ to obtain the solution to the reduced-rank regression problem; for this case, the appropriate $\boldsymbol{\mu}$ and $\mathbf{C}$ having reduced-rank t are given by

$$\mathbf{C} = \mathbf{\Sigma}_{YY}^{1/2}\left(\sum_{j=1}^{t} \mathbf{V}_j\mathbf{V}_j^\tau\right)\mathbf{\Sigma}_{YY}^{-1/2}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_Y - \mathbf{C}\boldsymbol{\mu}_X, \tag{4.6}$$

where $\mathbf{V}_j$ is the jth latent vector of the matrix $\mathbf{R}$ in (4.5). From $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{G}$ and $\mathbf{H}$ can be obtained in the following way. When $t < s$, $\mathbf{H}$ is derived from $\mathbf{A}$ by first computing $\mathbf{A}^+$ for the case $t = s$, and then taking only the first t rows of $\mathbf{A}^+$ for $\mathbf{H}$. The matrix $\mathbf{G}$ is the same as the matrix $\mathbf{B}$. Denoting the jth row of $\mathbf{G}$ and $\mathbf{H}$ by $\mathbf{g}_j^\tau$ and $\mathbf{h}_j^\tau$, respectively, we then define the jth *pair of canonical variates*, (ξ_j, ω_j), as $\xi_j = \mathbf{g}_j^\tau\mathbf{X}$ and $\omega_j = \mathbf{h}_j^\tau\mathbf{Y}$, $j = 1, 2,\ldots, t$. If we further assume that both $\mathbf{g}_j$ and $\mathbf{h}_j$ have unit length (i.e., $\mathbf{g}_j^\tau\mathbf{g}_j = 1$ and $\mathbf{h}_k^\tau\mathbf{h}_k = 1$ for $j = 1, 2,\ldots, r$ and $k = 1, 2,\ldots, s$) and set $\rho_j = \lambda_j^{1/2}$, for $j = 1, 2,\ldots, s$, then $\operatorname{corr}\{\xi_j, \xi_k\} = \delta_{jk}$, $(j, k = 1, 2,\ldots, r)$, $\operatorname{corr}\{\xi_j, \omega_k\} = \rho_j\delta_{jk}$, $(j = 1, 2,\ldots, r$, $k = 1, 2,\ldots, s)$, and $\operatorname{corr}\{\omega_j, \omega_k\} = \delta_{jk}$, $(j, k = 1, 2,\ldots, s)$, where δ_{jk} is the Kronecker delta. The ρ_j above is usually called the jth *canonical correlation coefficient.*

We shall henceforth refer to the model (2.3) together with the minimisation criterion (2.4), where $\mathbf{\Gamma} = \mathbf{\Sigma}_{YY}^{-1}$, as the *reduced-rank regression model corresponding to the canonical variates situation.* The solution (4.6) will be referred to similarly.

5. Sample Estimates

In general, the quantities which appear in the solution to the reduced-rank regression problem will be unknown and will have to be estimated using a sample of size n on the appropriate variates.

(a) For the *principal components* situation, let $\mathbf{X}_j$, $j = 1, 2,..., n$, be a sample of values on the r-vector $\mathbf{X}$, where, without loss of generality, we take $\boldsymbol{\mu}_X = \mathbf{0}$. A suitable estimator for the covariance matrix of $\mathbf{X}$ can then be given by $\hat{\boldsymbol{\Sigma}}_{XX} = n^{-1} \sum_{j=1}^{n} \mathbf{X}_j \mathbf{X}_j^{\tau}$. The regression coefficient matrix $\mathbf{C}$ having rank t in (3.5) is estimated by

$$\hat{\mathbf{C}} = \sum_{j=1}^{t} \hat{\mathbf{V}}_j \hat{\mathbf{V}}_j^{\tau}, \tag{5.1}$$

where $\hat{\mathbf{V}}_j$ is the latent vector corresponding to the jth largest latent root of $\hat{\boldsymbol{\Sigma}}_{XX}$.

(b) For the *canonical variates* case, let

$$\begin{bmatrix} \mathbf{X}_j \\ \mathbf{Y}_j \end{bmatrix}, \qquad j = 1, 2,..., n, \tag{5.2}$$

represent a sample of values on the variate (2.1), where we assume, without loss of generality, that $\boldsymbol{\mu}_X = \mathbf{0}$ and $\boldsymbol{\mu}_Y = \mathbf{0}$. We can then estimate the components of $\boldsymbol{\Sigma}$ in (2.2) by

$$\hat{\boldsymbol{\Sigma}}_{XX} = n^{-1} \sum_{j=1}^{n} \mathbf{X}_j \mathbf{X}_j^{\tau}, \quad \hat{\boldsymbol{\Sigma}}_{XY} = n^{-1} \sum_{j=1}^{n} \mathbf{X}_j \mathbf{Y}_j^{\tau} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_{YY} = n^{-1} \sum_{j=1}^{n} \mathbf{Y}_j \mathbf{Y}_j^{\tau}.$$

The reduced-rank regression coefficient matrix $\mathbf{C}$ with rank t in (4.6) is then estimated by

$$\hat{\mathbf{C}} = \hat{\boldsymbol{\Sigma}}_{YY}^{1/2} \left(\sum_{j=1}^{t} \hat{\mathbf{V}}_j \hat{\mathbf{V}}_j^{\tau} \right) \hat{\boldsymbol{\Sigma}}_{YY}^{-1/2} \hat{\boldsymbol{\Sigma}}_{YX} \hat{\boldsymbol{\Sigma}}_{XX}^{-1}, \tag{5.3}$$

where $\hat{\mathbf{V}}_j$ is the latent vector corresponding to the jth largest latent root of

$$\hat{\mathbf{R}} = \hat{\boldsymbol{\Sigma}}_{YY}^{-1/2} \hat{\boldsymbol{\Sigma}}_{YX} \hat{\boldsymbol{\Sigma}}_{XX}^{-1} \hat{\boldsymbol{\Sigma}}_{XY} \hat{\boldsymbol{\Sigma}}_{YY}^{-1/2}.$$

6. Perturbation Expansions

To obtain results concerning the asymptotic behaviour of the sample reduced-rank regression coefficient matrices (5.1) and (5.3) above, we use the technique known variously as the "delta-method" or "perturbation-method". These sample reduced-rank regression coefficient matrices depend on certain latent vectors: the $\hat{\mathbf{C}}$ of (5.1) depends on the latent vectors of $\hat{\mathbf{\Sigma}}_{XX}$, while the $\hat{\mathbf{C}}$ of (5.3) depends on those of $\hat{\mathbf{R}}$. To relate sample estimators, $\hat{\mathbf{C}}$, to their population counterparts, $\mathbf{C}$, it is, therefore, necessary to determine, first, the relationship between the latent vectors of $\hat{\mathbf{\Sigma}}_{XX}$ and $\mathbf{\Sigma}_{XX}$ and between those of $\hat{\mathbf{R}}$ and $\mathbf{R}$.

Writing $\hat{\mathbf{\Sigma}}_{XX} = \mathbf{\Sigma}_{XX} + \epsilon \mathbf{\Delta}_{XX}$ and similarly for $\hat{\mathbf{\Sigma}}_{XY}$ and $\hat{\mathbf{\Sigma}}_{YY}$, where the $\mathbf{\Delta}$'s represent the differences in approximation scaled by a factor ϵ to be thought of as real-valued and small, we get that $\hat{\mathbf{R}}$ is related to $\mathbf{R}$ by a power-series in ϵ. See (6.10) below. More generally, suppose $\mathbf{A}$ is a real symmetric matrix whose 'perturbation' $\mathbf{M}$ can be viewed as a power-series in ϵ:

$$\mathbf{M} = \mathbf{A} + \epsilon \mathbf{A}_1 + \epsilon^2 \mathbf{A}_2 + \cdots, \tag{6.1}$$

where we assume also that $\mathbf{M}$ is real symmetric. This perturbation of $\mathbf{A}$ induces corresponding perturbations in the latent roots $\{\lambda_j, j = 1, 2,..., J\}$ and vectors $\{\mathbf{V}_j, j = 1, 2,..., J\}$ of $\mathbf{A}$ which can be assumed to take the form,

$$\mu_j = \lambda_j + \epsilon \lambda_j^{(1)} + \epsilon^2 \lambda^{(2)} + \cdots, \tag{6.2}$$

$$\mathbf{W}_j = \mathbf{V}_j + \epsilon \mathbf{V}_j^{(1)} + \epsilon^2 \mathbf{V}_j^{(2)} + \cdots, \tag{6.3}$$

where μ_j and $\mathbf{W}_j$ are the jth latent root and vector, respectively, of $\mathbf{M}$, $j = 1, 2,..., J$. Since the $\{\mathbf{V}_j, j = 1, 2,..., J\}$ span E^J, then for all $l = 1, 2,...,$ the following relationship holds:

$$\mathbf{V}_j^{(l)} = \sum_{\substack{k=1 \\ k \neq j}}^{J} \alpha_{jk}^{(l)} \mathbf{V}_k, \tag{6.4}$$

for a set of coefficients $\{\alpha_{jk}^{(l)}\}$ to be determined (see [17]).

Equating the coefficients of appropriate powers of ϵ by substituting (6.2), (6.3) into the characteristic equation $\mathbf{M}\mathbf{W}_j = \mu_j \mathbf{W}_j$, we obtain the perturbation expansions for μ_j and $\mathbf{W}_j$ in terms of λ_j and $\mathbf{V}_j$. These are given in the following theorem.

THEOREM 3. *Let μ_j and $\mathbf{W}_j$, $j = 1, 2,..., J$, be the latent roots and vectors, respectively, of the real $(J \times J)$-symmetric matrix $\mathbf{M}$, and let λ_j and $\mathbf{V}_j$, $j = 1, 2,..., J$, be the corresponding latent roots and vectors respectively of the real $(J \times J)$-symmetric matrix $\mathbf{A}$, where $\mathbf{A}$ and $\mathbf{M}$ are related by (6.1). Suppose further that the first m latent roots of $\mathbf{A}$ are distinct, i.e., $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, and that the rest are zero, i.e., $\lambda_{m+1} = \cdots = \lambda_J = 0$.*

Then, for $j = 1, 2, \ldots, m$,

$$\mu_j = \lambda_j + \epsilon a_{jj}^{(1)} + \epsilon^2 \left[a_{jj}^{(2)} + \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} \{a_{jk}^{(1)}\}^2 \right]$$

$$+ \epsilon^3 \left[a_{jj}^{(3)} + \sum_{\substack{i=1 \\ i \neq j}}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{ji} \lambda_{jk} a_{ik}^{(1)} a_{kj}^{(1)} a_{ji}^{(1)} \right.$$

$$\left. - a_{jj}^{(1)} \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk}^2 \{a_{jk}^{(1)}\}^2 + 2 \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} a_{jk}^{(2)} a_{jk}^{(1)} \right] + O(\epsilon^4), \tag{6.5}$$

$$\mathbf{W}_j = \mathbf{V}_j + \epsilon \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} a_{jk}^{(1)} \mathbf{V}_k$$

$$+ \epsilon^2 \left[\sum_{\substack{i=1 \\ i \neq j}}^{J} \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{ji} \lambda_{jk} a_{ik}^{(1)} a_{jk}^{(1)} \mathbf{V}_i - a_{jj}^{(1)} \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} a_{jk}^{(1)} \mathbf{V}_k \right.$$

$$\left. + \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} a_{jk}^{(2)} \mathbf{V}_k \right] + O(\epsilon^3), \tag{6.6}$$

where $a_{jk}^{(l)} = \mathbf{V}_j^\tau \mathbf{A}_l \mathbf{V}_k = a_{kj}^{(l)}$, *and* $\lambda_{jk} = (\lambda_j - \lambda_k)^{-1}$, *for* $l = 1, 2, \ldots,$ *and* j, $k = 1, 2, \ldots, J$.

Compare (6.5) with a similar result obtained by Lawley ([7, Eq. (3)] for the roots of a covariance matrix. (An extension of Theorem 3 to the case in which $\mathbf{A}$ has q sets of equal roots is given in [5, Theor. 3.2.2].) Thus, if $\mathbf{A}$ and $\mathbf{M}$ satisfy the conditions of the theorem, then from (6.6) we have that

$$\mathbf{W}_j \mathbf{W}_j^\tau = \mathbf{V}_j \mathbf{V}_j^\tau + \epsilon \sum_{\substack{k=1 \\ k \neq j}}^{J} \lambda_{jk} a_{jk}^{(1)} \{\mathbf{V}_j \mathbf{V}_k^\tau + \mathbf{V}_k \mathbf{V}_j^\tau\} + O(\epsilon^2), \tag{6.7}$$

where $\mathbf{W}_j$ and $\mathbf{V}_j$, $j = 1, 2, \ldots, J$, are the latent vectors of $\mathbf{M}$ and $\mathbf{A}$, respectively, and $a_{jk}^{(1)}$ and λ_{jk} are as given in the theorem.

The perturbation approximations for both versions of $\hat{\mathbf{C}}$ (in terms of C) can now be given. They are each of the form

$$\hat{\mathbf{C}} = \mathbf{C} + \epsilon \sum_{j=1}^{t} \mathbf{\Delta}_j + O(\epsilon^2), \tag{6.8}$$

where $\mathbf{\Delta}_j$ depends on which $\mathbf{C}$ we are considering. The following will make this relationship more explicit.

(a) For the *principal components* situation, $\mathbf{C}$ is of the form (3.5) and $\hat{\mathbf{C}}$ is given by (5.1). Let λ_j, $\mathbf{V}_j$, $j = 1, 2,..., r$, be the latent roots (assumed distinct) and associated vectors of $\mathbf{\Sigma}_{XX}$, respectively. Set $\mathbf{E}_j = \mathbf{V}_j\mathbf{V}_j^\tau$, $j = 1, 2,..., r$, so that $\mathbf{C} = \sum_{j=1}^{t} \mathbf{E}_j$. Then the appropriate expansion is given by (6.8), where

$$\mathbf{\Delta}_j = \sum_{\substack{k=1\\k\neq j}}^{r} \lambda_{jk}\{\mathbf{E}_j\mathbf{\Delta}_{XX}\mathbf{E}_k + \mathbf{E}_k\mathbf{\Delta}_{XX}\mathbf{E}_j\} \tag{6.9}$$

and $\lambda_{jk} = (\lambda_j - \lambda_k)^{-1}$.

(b) The other case corresponds to the *canonical variates* situation, where $\mathbf{C}$ is of the form (4.6) and $\hat{\mathbf{C}}$ is given by (5.3). First, we need the following technical fact which can be obtained as a special case of results in Okamoto and Fujikoshi [12]. Let $\mathbf{\Sigma}_{YY} = \mathbf{Q\Lambda Q}^\tau$, where $\mathbf{Q}$ is orthogonal and $\mathbf{\Lambda} = \mathrm{diag}\{\nu_1, \nu_2,..., \nu_s\}$. Then,

$$\hat{\mathbf{\Sigma}}_{YY}^{1/2} = (\mathbf{\Sigma}_{YY} + \epsilon\mathbf{\Delta}_{YY})^{1/2} = \mathbf{\Sigma}_{YY}^{1/2} + \epsilon(\mathbf{QZQ}^\tau) + O(\epsilon^2),$$

$$\hat{\mathbf{\Sigma}}_{YY}^{-1/2} = (\mathbf{\Sigma}_{YY} + \epsilon\mathbf{\Delta}_{YY})^{-1/2} = \mathbf{\Sigma}_{YY}^{-1/2} - \epsilon[\mathbf{\Sigma}_{YY}^{-1/2}(\mathbf{QZQ}^\tau)\,\mathbf{\Sigma}_{YY}^{-1/2}] + O(\epsilon^2),$$

where

$$\mathbf{Z} = [z_{ij}], \qquad z_{ij} = \nu_{ij}(\mathbf{Q}^\tau\mathbf{\Delta}_{YY}\mathbf{Q})_{ij}, \qquad \nu_{ij} = (\nu_i^{1/2} + \nu_j^{1/2})^{-1}.$$

Now, let ρ_j^2, $\mathbf{V}_j$, $j = 1, 2,..., s$, be the latent roots (assumed distinct) and vectors of $\mathbf{R}$, respectively. Set $\mathbf{E}_j = \mathbf{V}_j\mathbf{V}_j^\tau$, $\mathbf{P}_j = \mathbf{\Sigma}_{YY}^{1/2}\mathbf{E}_j\mathbf{\Sigma}_{YY}^{-1/2}$, and $\mathbf{C}_j = \mathbf{P}_j\mathbf{\theta}$, $j = 1, 2,..., s$, where $\mathbf{\theta} = \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}$ is the full-rank regression coefficient matrix. Let $\mathbf{F} = \mathbf{\Sigma}_{YY}^{-1/2}\mathbf{\theta}$, so that $\mathbf{R} = \mathbf{F\Sigma}_{XX}\mathbf{F}^\tau$. Using the above results we obtain the perturbation expansion for $\hat{\mathbf{R}}$ in terms of $\mathbf{R}$, i.e.,

$$\hat{\mathbf{R}} = \mathbf{R} + \epsilon\mathbf{\Delta}_R + O(\epsilon^2), \tag{6.10}$$

where

$$\begin{aligned}\mathbf{\Delta}_R = {} & \{\mathbf{F\Delta}_{XY}\mathbf{\Sigma}_{YY}^{-1/2} + \mathbf{\Sigma}_{YY}^{-1/2}\mathbf{\Delta}_{YX}\mathbf{F}^\tau\} - \mathbf{F\Delta}_{XX}\mathbf{F}^\tau \\ & - \{\mathbf{\Sigma}_{YY}^{-1/2}(\mathbf{QZQ}^\tau)\mathbf{R} + \mathbf{R}(\mathbf{QZQ}^\tau)\,\mathbf{\Sigma}_{YY}^{-1/2}\}.\end{aligned} \tag{6.11}$$

The appropriate perturbation expansion for $\hat{\mathbf{C}}$ in (5.3) in terms of $\mathbf{C}$ in (4.6) is therefore given by (6.8), where this time,

$$\begin{aligned}\mathbf{\Delta}_j = {} & \sum_{\substack{k=1\\k\neq j}}^{s} \rho_{jk}\mathbf{\Sigma}_{YY}^{1/2}\{\mathbf{E}_j\mathbf{\Delta}_R\mathbf{E}_k + \mathbf{E}_k\mathbf{\Delta}_R\mathbf{E}_j\}\mathbf{F} + \mathbf{P}_j\{\mathbf{\Delta}_{YX} - \mathbf{\theta\Delta}_{XX}\}\,\mathbf{\Sigma}_{XX}^{-1} \\ & - \{\mathbf{P}_j(\mathbf{QZQ}^\tau) - (\mathbf{QZQ}^\tau)\,\mathbf{E}_j\}\mathbf{F},\end{aligned} \tag{6.12}$$

and $\rho_{jk} = (\rho_j^2 - \rho_k^2)^{-1}$.

7. Asymptotic Distribution Theory

In addition to the perturbation expansions of the previous section, we need the following results. Henceforth, vec $\mathbf{A}$ will denote the $(mn \times 1)$-column vector formed by placing the columns of the $(m \times n)$-matrix $\mathbf{A}$ under one another successively, and the Kronecker Product of the $(m \times n)$-matrix $\mathbf{A}$ with the $(p \times q)$-matrix $\mathbf{B}$ will be defined as the block matrix $\mathbf{A} \otimes \mathbf{B} = [\mathbf{A}b_{jk}]$, having mp rows and nq columns. They are related by the fact that $\operatorname{vec}(\mathbf{ABC}) = (\mathbf{A} \otimes \mathbf{C}^\tau) \operatorname{vec} \mathbf{B}$. The Permuted Identity Matrix, $\mathbf{I}_{(m,n)}$, will denote an $(mn \times mn)$-matrix partitioned into $(m \times n)$-submatrices such that the ijth-submatrix has a 1 in its jith position and zeros elsewhere (see [8]). If $\mathbf{A}$ is $(m \times n)$ and $\mathbf{B}$ is $(p \times q)$, then $\mathbf{I}_{(p,m)}(\mathbf{B} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{B})\mathbf{I}_{(q,n)}$. We shall also need the following two well-known results; the symbol $\to^{\mathscr{L}}$ is used to denote convergence in distribution.

Lemma 1. *Let* $n\mathbf{S} \sim W_p(n, \mathbf{\Sigma})$. *Then, as* $n \to \infty$, $n^{1/2}\operatorname{vec}\{\mathbf{S} - \mathbf{\Sigma}\} \to^{\mathscr{L}} N_{p^2}(\mathbf{0}, \mathbf{Z}_p)$, *where*

$$\mathbf{Z}_p = (\mathbf{I}_{p^2} + \mathbf{I}_{(p,p)})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) = (\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \mathbf{I}_{(p,p)}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}). \tag{7.1}$$

Lemma 2. *Let* $\mathbf{x}_n$ *be a sequence of* p*-dimensional random vectors, and suppose that as* $n \to \infty$, $n^{1/2}(\mathbf{x}_n - \boldsymbol{\mu}) \to^{\mathscr{L}} N_p(\mathbf{0}, \mathbf{\Psi})$, *where* $\mathbf{\Psi}$ *is a* $(p \times p)$*-matrix of finite constants. Let* $\boldsymbol{\omega} = \boldsymbol{\omega}(\mathbf{x})$ *be a* q*-dimensional vector whose entries have continuous derivatives in a neighborhood of* $\mathbf{x} = \boldsymbol{\mu}$. *Then, as* $n \to \infty$, $n^{1/2}(\boldsymbol{\omega}(\mathbf{x}_n) - \boldsymbol{\omega}(\boldsymbol{\mu})) \to^{\mathscr{L}} N_q(\mathbf{0}, \boldsymbol{\xi}^\tau\mathbf{\Psi}\boldsymbol{\xi})$, *where* $\boldsymbol{\xi} = (\partial\boldsymbol{\omega}(\mathbf{x})/\partial\mathbf{x})_{\mathbf{x}=\boldsymbol{\mu}}$ *is the* $(p \times q)$*-matrix of first-order partials of* $\boldsymbol{\omega}$ *evaluated at* $\boldsymbol{\mu}$.

Finally, we need results on matrix differentiation. A convenient notation has been developed by Neudecker [9, 10]. However, since our definition of a Kronecker product is the reverse of his, the results are rearranged accordingly. Let $\boldsymbol{\phi}(\mathbf{A})$ be an $(m \times n)$-matrix-valued function of an $(s \times t)$-matrix $\mathbf{A}$, and suppose that $d\boldsymbol{\phi}(\mathbf{A}) = \sum_k \mathbf{M}_k(d\mathbf{A})\mathbf{N}_k$, $\mathbf{M}_k$ and $\mathbf{N}_k$ being conformable matrices. Then, $\partial \operatorname{vec} \boldsymbol{\phi}(\mathbf{A})/\partial \operatorname{vec} \mathbf{A} = \sum_k (\mathbf{M}_k^\tau \otimes \mathbf{N}_k)$. If $\mathbf{A}$ is partitioned into the form

$$\mathbf{A} = \begin{array}{cc} \begin{array}{cc} t_1 & t_2 \end{array} & \\ \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] & \begin{array}{c} s_1 \\ s_2 \end{array} \end{array}, \tag{7.2}$$

then the matrix of first-order partials of $\phi(\mathbf{A})$ with respect to $\mathbf{A}$ is given by the $(st \times mn)$-matrix

$$\partial \operatorname{vec} \phi(\mathbf{A})/\partial \operatorname{vec} \mathbf{A} = [\xi_{11}^{\tau} \mid \xi_{21}^{\tau} \mid \xi_{12}^{\tau} \mid \xi_{22}^{\tau}]^{\tau}, \tag{7.3}$$

where ξ_{jk} is the $(s_j t_k \times mn)$-matrix $\partial \operatorname{vec} \phi(\mathbf{A})/\partial \operatorname{vec} \mathbf{A}_{jk}$.

We now use the perturbation expansions of Section 6 together with the above to obtain the asymptotic distributions of the various reduced-rank regression coefficient matrices. To do this, we neglect terms of order ϵ^2 and above and replace $\epsilon\mathbf{\Delta}_{XX}$ by $d\hat{\mathbf{\Sigma}}_{XX}$, $\epsilon\mathbf{\Delta}_{XY}$ by $d\hat{\mathbf{\Sigma}}_{XY}$, and $\epsilon\mathbf{\Delta}_{YY}$ by $d\hat{\mathbf{\Sigma}}_{YY}$. This now permits us to use the results on matrix differentiation.

(a) For the *principal components* situation, the sample reduced-rank regression coefficient matrix $\hat{\mathbf{C}}$ having rank t and given by (5.1) with perturbation expansion (6.8), (6.9) has derivative

$$\xi_{XX}^{(C)} = \sum_{j=1}^{t} \sum_{\substack{k=1\\k\neq j}}^{r} \lambda_{jk}\{(\mathbf{E}_j \otimes \mathbf{E}_k) + (\mathbf{E}_k \otimes \mathbf{E}_j)\}. \tag{7.4}$$

By using the relationship, $\mathbf{\Sigma}_{XX}\mathbf{V}_j = \lambda_j\mathbf{V}_j$, we get that

$$\begin{aligned}
&\xi_{XX}^{(C)\tau}(\mathbf{\Sigma}_{XX} \otimes \mathbf{\Sigma}_{XX})\, \xi_{XX}^{(C)}\\
&\quad= \sum_{j=1}^{t} \sum_{l=1}^{t} \sum_{\substack{k=1\\k\neq j}}^{r} \sum_{\substack{m=1\\m\neq l}}^{r} \lambda_{jk}\lambda_{lm}\lambda_j\lambda_k\{(\mathbf{E}_j\mathbf{E}_l \otimes \mathbf{E}_k\mathbf{E}_m) + (\mathbf{E}_j\mathbf{E}_m \otimes \mathbf{E}_k\mathbf{E}_l)\\
&\qquad + (\mathbf{E}_k\mathbf{E}_l \otimes \mathbf{E}_j\mathbf{E}_m) + (\mathbf{E}_k\mathbf{E}_m \otimes \mathbf{E}_j\mathbf{E}_l)\}.
\end{aligned} \tag{7.5}$$

If we denote the *summand* in (7.5) by the 4-tuple (j, k, l, m), then

$$(j, k, j, k) = \lambda_{jk}^2\lambda_j\lambda_k\{(\mathbf{E}_j \otimes \mathbf{E}_k) + (\mathbf{E}_k \otimes \mathbf{E}_j)\} = -(j, k, k, j), \tag{7.6}$$

and

$$(j, k, l, m) = 0 \qquad \text{for} \quad \begin{cases} j \neq l & \text{or} \quad k \neq m,\\ j \neq m & \text{or} \quad k \neq l,\\ j \neq k \neq l \neq m. \end{cases} \tag{7.7}$$

It is now relatively straightforward to compute the asymptotic covariance matrix of $\hat{\mathbf{C}}$ in (5.1). We summarize as follows.

THEOREM 4. *Let $\mathbf{X}_j$, $j = 1, 2,\ldots, n$, be a random sample of values from the distribution, $N_r(\mathbf{0}, \mathbf{\Sigma}_{XX})$. Let λ_j and $\mathbf{V}_j$, $j = 1, 2,\ldots, r$, be the latent roots and vectors respectively of $\mathbf{\Sigma}_{XX}$. Suppose the λ_j are distinct. Set $\mathbf{E}_j = \mathbf{V}_j\mathbf{V}_j^{\tau}$, $j = 1, 2,\ldots, r$. If $\hat{\mathbf{C}}$ in* (5.1) *is the estimator of $\mathbf{C}$ in* (3.5), *then as $n \to \infty$, $n^{1/2} \operatorname{vec}\{\hat{\mathbf{C}} - \mathbf{C}\}$ is* *asymptotically distributed as an r^2-variate normal with mean zero and covariance matrix*

$$(\mathbf{I}_{r^2} + \mathbf{I}_{(r,r)}) \sum_{j=1}^{t} \sum_{k=t+1}^{r} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2} \{(\mathbf{E}_j \otimes \mathbf{E}_k) + (\mathbf{E}_k \otimes \mathbf{E}_j)\}. \tag{7.8}$$

(b) We now obtain the asymptotic distribution of the reduced-rank regression coefficient matrix for the *canonical variates* case, namely when $\mathbf{C}$ is given by (4.6) and is estimated by $\hat{\mathbf{C}}$ in (5.3). From (6.11), (6.12), we obtain the following derivatives:

$$\xi_{XX}^{(C)} = -(\mathbf{C}^\tau \otimes \mathbf{\Sigma}_{XX}^{-1}) - \sum_{j=1}^{t} \sum_{\substack{k=1 \\ k \neq j}}^{s} \rho_{jk} \{(\mathbf{C}_j^{\tau} \otimes \mathbf{F}^\tau \mathbf{E}_k \mathbf{F}) + (\mathbf{C}_k^{\tau} \otimes \mathbf{F}^\tau \mathbf{E}_j \mathbf{F})\}, \tag{7.9}$$

$$\xi_{YX}^{(C)} = \left[\left(\sum_{j=1}^{t} \mathbf{P}_j^{\tau} \right) \otimes \mathbf{\Sigma}_{XX}^{-1} \right] + \sum_{j=1}^{t} \sum_{k=1}^{s} \rho_{jk} \{(\mathbf{P}_j^{\tau} \otimes \mathbf{F}^\tau \mathbf{E}_k \mathbf{F}) + (\mathbf{P}_k^{\tau} \otimes \mathbf{F}^\tau \mathbf{E}_j \mathbf{F})\}, \tag{7.10}$$

$$\xi_{XY}^{(C)} = \sum_{j=1}^{t} \sum_{\substack{k=1 \\ k \neq j}}^{s} \rho_{jk} \{(\mathbf{C}_j^{\tau} \otimes \mathbf{\Sigma}_{YY}^{-1} \mathbf{C}_k) + (\mathbf{C}_k^{\tau} \otimes \mathbf{\Sigma}_{YY}^{-1} \mathbf{C}_j)\}, \tag{7.11}$$

$$\xi_{YY}^{(C)} = (\mathbf{Q} \otimes \mathbf{Q})^\tau \left[\sum_{j=1}^{t} \mathbf{M}_j \right] \mathbf{\Lambda}^* (\mathbf{Q} \otimes \mathbf{Q}), \tag{7.12}$$

where

$$\mathbf{\Lambda}^* = \operatorname{diag}\{\nu_{11}, \nu_{21}, \ldots, \nu_{s1}, \nu_{12}, \nu_{22}, \ldots, \nu_{s2}, \ldots, \nu_{1s}, \nu_{2s}, \ldots, \nu_{ss}\}, \tag{7.13}$$

$$\begin{aligned} \mathbf{M}_j = \Big[&\{(\mathbf{I} \otimes \mathbf{E}_j \mathbf{F}) - (\mathbf{P}_j^{\tau} \otimes \mathbf{F})\} \\ &- \sum_{\substack{k=1 \\ k \neq j}}^{s} \rho_{jk} \{\rho_j^2 (\mathbf{P}_j^{\tau} \otimes \mathbf{E}_k \mathbf{F}) + \rho_k^2 (\mathbf{E}_k \mathbf{\Sigma}_{YY}^{1/2} \otimes \mathbf{\Sigma}_{YY}^{-1} \mathbf{C}_k) \\ &+ \rho_j^2 (\mathbf{\Sigma}_{YY}^{-1/2} \mathbf{E}_k \otimes \mathbf{E}_j \mathbf{F}) + \rho_k^2 (\mathbf{E}_k \otimes \mathbf{\Sigma}_{YY}^{-1} \mathbf{C}_j)\} \Big], \end{aligned} \tag{7.14}$$

where we have used the relationship, $\mathbf{RV}_j = \rho_j^2 \mathbf{V}_j$, $j = 1, 2, \ldots, s$, and where $\rho_{jk} = (\rho_j^2 - \rho_k^2)^{-1}$.

Now define the (partitioned) matrix,

$$\xi^{(C)} = [\xi_{XX}^{(C)\tau} \mid \xi_{YX}^{(C)\tau} \mid \xi_{XY}^{(C)\tau} \mid \xi_{YY}^{(C)\tau}]^\tau = \left(\frac{\partial \operatorname{vec} \hat{\mathbf{C}}}{\partial \operatorname{vec} \hat{\mathbf{\Sigma}}} \right)_{\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}}. \tag{7.15}$$

From Lemmas 1 and 2, we have the following theorem.

THEOREM 5. *Let* (5.2) *be a random sample from the distribution*

$$N_{r+s}(\mathbf{0}, \mathbf{\Sigma}), \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix}, \tag{7.16}$$

and let $\hat{\mathbf{C}}$ *in* (5.3) *be an estimator of* $\mathbf{C}$ *in* (4.6), *where we assume that the latent roots of* $\mathbf{R}$ *in* (4.5) *are distinct. Then, as* $n \to \infty$, $n^{1/2} \operatorname{vec}\{\hat{\mathbf{C}} - \mathbf{C}\}$ *is asymptotically distributed as an sr-variate normal with mean zero and covariance matrix*

$$\boldsymbol{\xi}^{(C)\tau}(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\boldsymbol{\xi}^{(C)} + \boldsymbol{\xi}^{(C)\tau}\mathbf{I}_{(r+s,r+s)}(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\boldsymbol{\xi}^{(C)}, \tag{7.17}$$

where $\boldsymbol{\xi}^{(C)}$ *is given by* (7.15).

Unfortunately, because of the complicated nature of Eqs. (7.9)–(7.14), the covariance matrix (7.17) does not seem to be expressible in a more explicit fashion as was (7.8).

(c) We now have the following well-known result which is a corollary of Theorem 5, and concerns *full-rank regression*, i.e., when $t = s$. The full-rank regression coefficient matrix is $\boldsymbol{\theta} = \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}$ and is estimated by $\hat{\boldsymbol{\theta}} = \hat{\mathbf{\Sigma}}_{YX}\hat{\mathbf{\Sigma}}_{XX}^{-1}$. The matrix of first-order partials, $\boldsymbol{\xi}^{(C)}$, given by (7.15), reduces to

$$\begin{aligned} \boldsymbol{\xi}^{(C)} &= [-\boldsymbol{\theta} \otimes \mathbf{\Sigma}_{XX}^{-1} \,\vdots\, \mathbf{I}_s \otimes \mathbf{\Sigma}_{XX}^{-1} \,\vdots\, \mathbf{0} \,\vdots\, \mathbf{0}]^\tau, \\ &= ([-\boldsymbol{\theta} \,\vdots\, \mathbf{I}_s] \otimes [\mathbf{\Sigma}_{XX}^{-1} \,\vdots\, \mathbf{0}])^\tau, \end{aligned}$$

whence

$$\begin{aligned} \boldsymbol{\xi}^{(C)\tau}\mathbf{I}_{(r+s,r+s)} &= \mathbf{I}_{(sr,sr)}([\mathbf{\Sigma}_{XX}^{-1} \,\vdots\, \mathbf{0}] \otimes [-\boldsymbol{\theta} \,\vdots\, \mathbf{I}_s]), \\ &= \mathbf{I}_{(sr,sr)}[-\mathbf{\Sigma}_{XX}^{-1} \otimes \boldsymbol{\theta} \,\vdots\, \mathbf{0} \,\vdots\, \mathbf{\Sigma}_{XX}^{-1} \otimes \mathbf{I}_s \,\vdots\, \mathbf{0}], \end{aligned}$$

so that

$$\boldsymbol{\xi}^{(C)\tau}(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\boldsymbol{\xi}^{(C)} = (\mathbf{\Sigma}_{YY} - \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}) \otimes \mathbf{\Sigma}_{XX}^{-1},$$

$$\boldsymbol{\xi}^{(C)\tau}\mathbf{I}_{(r+s,r+s)}(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\, \boldsymbol{\xi}^{(C)} = \mathbf{0}.$$

This gives the following result.

THEOREM 6. *Let* (5.2) *be a random sample from the distribution* (7.16), *and suppose* $\hat{\boldsymbol{\theta}}$ *is the estimator of the full-rank regression coefficient matrix* $\boldsymbol{\theta}$. *Then, as* $n \to \infty$, $n^{1/2} \operatorname{vec}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\}$ *is asymptotically distributed as an sr-variate normal with mean zero and covariance matrix* $\mathbf{\Sigma}_{\epsilon\epsilon} \otimes \mathbf{\Sigma}_{XX}^{-1}$, *where* $\mathbf{\Sigma}_{\epsilon\epsilon} = \mathbf{\Sigma}_{YY} - \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}$.

8. Approximate Confidence Bounds

Having found the form of the asymptotic distribution of vec $\hat{\mathbf{C}}$, we are now in a position to set (large-sample) confidence bounds on the unknown parameter, vec $\mathbf{C}$.

Denote the asymptotic covariance matrix of vec $\hat{\mathbf{C}}$ by the general symbol, $\boldsymbol{\Psi}_C$; for example, $n\boldsymbol{\Psi}_C$ would be (7.8) for the principal components case, and (7.17) for the canonical variates case, and in general, $\boldsymbol{\Psi}_C$ would be $O(n^{-1})$. An approximate $100(1-\gamma)$ percent confidence bound for α^τ vec $\mathbf{C}$, where α is an sr-vector of unit length, can then be given by

$$\boldsymbol{\alpha}^\tau \operatorname{vec} \hat{\mathbf{C}} \pm t_{n-r}^{\gamma}(\boldsymbol{\alpha}^\tau \hat{\boldsymbol{\Psi}}_C \boldsymbol{\alpha})^{1/2}, \tag{8.1}$$

where $\hat{\boldsymbol{\Psi}}_C$ is a consistent estimator of $\boldsymbol{\Psi}_C$, obtained perhaps by replacing population matrices by their respective sample estimates, and t_ν^γ is the $100(1-\gamma/2)$ percentage-point of the Student's t-distribution with ν degrees of freedom. Notice that $\boldsymbol{\Psi}_C$ depends on the rank of $\mathbf{C}$.

Suppose then, for purposes of illustration, that we wish to compute an appropriate confidence bound for the entries of the matrix $\mathbf{C}$ in the principal components case. For simplicity, we shall take $r = 2$ and $t = 1$. In this case, the full-rank regression coefficient matrix is $\mathbf{I}_2$, the (2×2)-identity matrix. Then, from (7.8), the appropriate form of (8.1) is given by

$$\boldsymbol{\alpha}^\tau \operatorname{vec} \hat{\mathbf{C}} \pm t_{n-2}^{\gamma} \left[\frac{\hat{\lambda}_1\hat{\lambda}_2}{n(\hat{\lambda}_1 - \hat{\lambda}_2)^2}\right]^{1/2} s_\alpha, \tag{8.2}$$

where

$$s_\alpha = [\boldsymbol{\alpha}^\tau(\mathbf{I}_4 + \mathbf{I}_{(2,2)}) \{(\hat{\mathbf{C}} \otimes (\mathbf{I}_2 - \hat{\mathbf{C}})) + ((\mathbf{I}_2 - \hat{\mathbf{C}}) \otimes \hat{\mathbf{C}})\}\boldsymbol{\alpha}]^{1/2}, \tag{8.3}$$

$\hat{\lambda}_1$, $\hat{\lambda}_2$ are the latent roots of $\hat{\boldsymbol{\Sigma}}_{XX}$, and $\boldsymbol{\alpha}$ is any 4-dimensional vector having unit length. If the vector $\boldsymbol{\alpha}$ has a 1 in the jth position and zeroes elsewhere, then we have the following approximate $100(1-\gamma)$ percent confidence intervals for the elements of the matrix $\mathbf{C} = [c_{jk}]$:

$$c_{11}\colon \hat{c}_{11} \pm 2t_{n-2}^{\gamma} \left[\frac{\hat{\lambda}_1\hat{\lambda}_2\hat{c}_{11}(1-\hat{c}_{11})}{n(\hat{\lambda}_1 - \hat{\lambda}_2)^2}\right]^{1/2};$$

$$c_{12}\colon \hat{c}_{12} \pm t_{n-2}^{\gamma} \left[\frac{\hat{\lambda}_1\hat{\lambda}_2}{n(\hat{\lambda}_1 - \hat{\lambda}_2)^2} \{(\hat{c}_{11} + \hat{c}_{22}) - 2(\hat{c}_{11}\hat{c}_{22} + \hat{c}_{21}\hat{c}_{12})\}\right]^{1/2};$$

$$c_{22}\colon \hat{c}_{22} \pm 2t_{n-2}^{\gamma} \left[\frac{\hat{\lambda}_1\hat{\lambda}_2\hat{c}_{22}(1-\hat{c}_{22})}{n(\hat{\lambda}_1 - \hat{\lambda}_2)^2}\right]^{1/2}.$$

The confidence interval for c_{21} is the same as that for c_{12} since $\mathbf{C}$ is symmetric. The extension to general r and t through (7.8) is immediate.

For the case of canonical variates, however, the expression for $\boldsymbol{\Psi}_C$ turns out to be more complicated, and consequently, (8.1) would be difficult to compute directly simply by replacing the unknown parameter values by their corresponding sample estimates. We, therefore, suggest using a technique known commonly as the "jackknife" to get another consistent estimator of $\boldsymbol{\Psi}_C$ which is more amenable to computation. For clarity in exposition, we omit all mention of the rank of the regression coefficient matrix which we wish to jackknife.

Let $\hat{\mathbf{C}}$ be the estimator (5.3) of $\mathbf{C}$ based on all n observations. Suppose the sample can be partitioned into m distinct groups, each group consisting of k observations (so that $n = mk$). Let $\hat{\mathbf{C}}_{(j)}$ be a version of $\hat{\mathbf{C}}$ obtained by deleting the jth group and using the remaining $n - k = (m - 1)k$ observations. Compute the *pseudovalues*, $\hat{\mathbf{C}}_{pj} = m\hat{\mathbf{C}} - (m - 1)\hat{\mathbf{C}}_{(j)}$. Do this m times. The jackknifed estimator of $\mathbf{C}$ is then the average of these pseudovalues, namely, $\hat{\mathbf{C}}_p = m^{-1} \sum_{j=1}^{m} \hat{\mathbf{C}}_{pj}$. As well as possessing the well-known property of reducing the asymptotic bias of $\hat{\mathbf{C}}$ from $O(n^{-1})$ to $O(n^{-2})$ (see [12]), $\hat{\mathbf{C}}_p$ also provides the basis for a convenient estimator of $\boldsymbol{\Psi}_C$ for the canonical variates case in place of the more complicated version derived in Section 7. Assuming the values $\hat{\mathbf{C}}_{pj}$ to be iid (which in general they are not), then

$$\mathbf{S}_p = \frac{1}{m(m-1)} \sum_{j=1}^{m} \{\mathrm{vec}(\hat{\mathbf{C}}_{pj} - \hat{\mathbf{C}}_p)\}\{\mathrm{vec}(\hat{\mathbf{C}}_{pj} - \hat{\mathbf{C}}_p)\}^{\tau} \tag{8.4}$$

can be used as an estimator of the asymptotic covariance matrix of $\mathrm{vec}\,\hat{\mathbf{C}}$ or $\mathrm{vec}\,\hat{\mathbf{C}}_p$. For any sr-vector $\boldsymbol{\alpha}$ of unit length, $\boldsymbol{\alpha}^{\tau}\{\mathrm{vec}(\hat{\mathbf{C}}_p - \mathbf{C})\}/(\boldsymbol{\alpha}^{\tau}\mathbf{S}_p\boldsymbol{\alpha})^{1/2}$ is distributed approximately as a Student's t-variate with $(m - 1)$ degrees-of-freedom. The corresponding approximate $100(1 - \gamma)$ percent confidence bound on $\boldsymbol{\alpha}^{\tau}\,\mathrm{vec}\,\mathbf{C}$ is, therefore, given by

$$\boldsymbol{\alpha}^{\tau}\ \mathrm{vec}\,\hat{\mathbf{C}}_p \pm t_{m-1}^{\gamma}(\boldsymbol{\alpha}^{\tau}\mathbf{S}_p\boldsymbol{\alpha})^{1/2}, \tag{8.5}$$

where t_{ν}^{γ} is the $100(1 - \gamma/2)$ percentage-point of the Student's t-distribution with ν degrees of freedom. Similar bounds can be constructed by replacing $\hat{\mathbf{C}}_p$ in (8.4) and (8.5) by $\hat{\mathbf{C}}$.

ACKNOWLEDGMENTS

This paper forms part of the author's doctoral dissertation at the University of California, Berkeley, written under the supervision and guidance of Professor David R. Brillinger. Conversations with Professor Robert Berk were also helpful in the preparation of this paper. The author also wishes to thank the referee for a critical reading of the paper and for the reference [12]. The investigation was partially supported by USPHS Research Grant No. GM-10525-09, National Institute of Health, Public Health Service.

References

[1] Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22** 327–351.

[2] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[3] Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

[4] Fujikoshi, Y. (1974). The Likelihood Ratio Tests for the Dimensionality of Regression Coefficients. *J. Multivariate Anal.* **4** 327–340.

[5] Izenman, A. J. (1972). Reduced-rank regression for the multivariate linear model, its relationship to certain classical multivariate techniques, and its application to the analysis of multivariate data. Ph.D. dissertation, University of California, Berkeley.

[6] Izenman, A. J. (1974). Residual Analysis for Reduced-Rank Regression and Canonical Variates. Technical Report No. 20, Dept. Statistics. Tel-Aviv University, Israel.

[7] Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43** 128–136.

[8] MacRae, E. C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Ann. Statist.* **2** 337–346.

[9] Neudecker, H. (1968). The Kronecker matrix product and some of its applications. *Statistica Neerlandica* **22** 69–82.

[10] Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Statist. Assoc.* **64** 953–963.

[11] Okamoto, M. (1969). Optimality of principal components. In *Multivariate Analysis-*II, (P. R. Krishnaiah, Ed.), pp. 673–685. Academic Press, New York.

[12] Okamoto, M., and Fujikoshi, Y. (1974). Perturbation of a matrix function and its application to multivariate analysis. Submitted for publication.

[13] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.

[14] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[15] Rao, C. R., and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

[16] Robinson, P. M. (1973). Generalized canonical analysis for time series. *J. Multivariate Anal.* **3** 140–160.

[17] Wilkinson, J. (1965). *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford.

[18] Krishnaiah, P. R. and Schuurmann, F. J. (1974). On the evaluation of some distributions that arise in simultaneous tests for the equality of the latent roots of the covariance matrix. *J. Multivariate Anal.* **4** 265–282.